



H20/B09 次世代ペタバイト情報ストレージシステムの研究(1節 共同プロジェクト研究の理念と概要, 第4章 共同プロジェクト研究)

雑誌名	東北大学電気通信研究所研究活動報告
巻	15
ページ	254-256
発行年	2009-08
URL	http://hdl.handle.net/10097/48438

次世代ペタバイト情報ストレージシステムの研究

〔1〕組織

代表者： 村岡 裕明

(東北大学電気通信研究所)

対応者： 村岡 裕明

(東北大学電気通信研究所)

分担者：

藤本和久：東北大学電気通信研究所：教授

サイモン・グリーブス：東北大学電気通信研究所：
准教授

青井基：東北大学電気通信研究所：客員教授

三浦健司：東北大学電気通信研究所：助手

青井基：東北大学電気通信研究所：教授

松岡浩：東北大学電気通信研究所・客員教授：

谷啓二：日本原子力研究開発機構那珂核融合研究
所：研究主席

山本康友：日立製作所システム開発研究所：部長

藤林昭：日立製作所システム開発研究所：主任研究
員

島田健太郎：日立製作所システム開発研究所：シニ
アアーキテクト

赤池洋俊：日立製作所システム開発研究所：研究員

小河卓二：日立製作所 RAID システム事業部開発本
部：RAID テクニカルスペシャリスト

浜田憲男：日立製作所 RAID システム事業部：部長
大沢寿：愛媛大学：教授

岡本好弘：愛媛大学：准教授

仲村泰明：愛媛大学：助教

金井靖：新潟工科大学：教授

山川清志：東北大学電気通信研究所：客員准教授

伊勢和幸：秋田県高度技術研究所：主任研究員

木谷貴則：秋田県高度技術研究所：研究員

研究費： 物件費 0 円 旅費 32 万 9 千円

〔2〕研究経過

本共同プロジェクトは、超大容量情報ストレージ技術に関する研究を行うことを目的にしている。人類が取り扱うデジタル情報は飛躍的な拡大を続けており、すでに 10 エクサバイト (10^{19} バイト) を超えているとの推定もある。通常の個人のパソコンにおけるストレージ容量が 100 GB (10^{11} バイト) 程度であることを考えると極めて大きな情報量であるこ

とが理解されるが、最近のネットワークコンセプトに基づくクラウドコンピューティングが急速な発展を続けているが、ここでは動画映像や長時間音楽情報配信など大規模データが大量に送受されており、情報ストレージの需要はさらに急速に伸びている。

この膨大な情報を電子的に蓄積してコンピュータで容易かつ効率的な利用を提供することが情報ストレージ技術の大きな目標である。インターネットなど大量の情報が頻繁に伝送されているが、これらは主としてハードディスク装置 (HDD)、さらには磁気テープ装置などの主として磁気を用いた情報ストレージからアクセスされていることが多い。

この際に、同時に、大容量情報のアクセスには高速のデータ転送能力がなければ、巨大な情報量を現実的な時間で処理できないので高速性も極めて重要である。現在の大規模情報ストレージサイトではテラバイトから数十テラバイト級のストレージシステムが主流であるが、急速な情報量の拡大を背景に近い将来にはペタバイト (10^{15} バイト) 級の容量が必要である。しかし、ストレージ容量を大きくするために、単に HDD のユニット数を比例的に増加させることはその大きさや消費電力などの観点から現実的ではなく、将来の具体的なシステム構成指針は必ずしも定まっていない。

このような背景から、ストレージ装置の高密度化とともにその高密度性を生かす用途を含めた総合的なストレージ研究を展開し、次世代ペタバイトストレージのシステム的な課題を抽出しその解決のアプローチを明らかにする研究を行うものである。

本共同研究は、大規模ストレージサブシステムの研究を目標に開始した昨年度の成果を受け、今年度は HDD 自体の高密度・大容量化も含めた総合的な研究に展開することとした。昨年度は多数の HDD を並列化した RAID 型ストレージサブシステムの省電力アーキテクチャを中心に議論を進めたが、HDD の高密度化と大容量化は省電力化にも直接貢献するものであり、このための研究も包括的に進めた。以下、研究活動状況の概要を示す。

研究打合せの開催状況は下記の通りである。

○ストレージサブシステムグループ (7 回)

2008 年 4 月 9 日

2008年6月3日
 2008年8月5日
 2008年10月22日
 2008年12月3日
 2009年2月5日
 2009年3月19日

○高密度磁気記録グループ (7回)

2008年6月4日
 2008年8月7日
 2008年9月9日
 2008年11月25日
 2008年12月19日
 2009年2月3日
 2009年3月13日

〔3〕 成果

(3-1) 研究成果

【ストレージサブシステムの成果】

これまでのストレージシステムではデータ信頼性や大容量性や高速性が重視され、その省電力性は問題にされなかった。このため、高速ディスク回転数の HDD を多数用いてシステムを構成してきた。ところが、最近の大容量動画情報の急速なネットワーク上での普及やクラウドコンピューティング化による大量の情報複製によってストレージ容量は年率 160% (5 年で 10 倍) とも言われる急速な増加を続けており、データセンターでのその消費電力は温室効果ガス排出や電力コスト、さらには廃熱や騒音・設置スペースなど今後ますます深刻化する大きな環境問題として注目されるに至った。(IEEE Spectrum, February 2009)

本共同研究ではすでに平成 17 年度ころより、新たな並列化ストレージシステムの方式提案を行って検討を続けてきた。昨年より検討してきた階層型大容量ストレージシステムの省電力化についてすでに提案方式を絞り込んでいたが、これを具体的な方式設計とその開発要素の抽出を行った。

ハードディスク装置(HDD)は磁気を用いる記録であるためにその記憶が不揮発性(電源を切っても記録情報を失わない)である。これはいったん記録された情報を保持するのに本質的にはまったく電力を必要としないことを意味している。これまでの方式では高速のデータ入出力を担う階層に消費電力の大きい高速 HDD を、その下位に使用頻度の下がったデータを対比させる低消費電力 HDD を置く階層化までは既知であったが、本グループではすべての情報を下位(ニアライン) HDD に蓄えた後にいったん電源を切り、使用される情報を予知して少量の高速 HDD にステージングする従来とは逆の発想に

よるストレージシステムの実現を目指してきた。このシステムは大きな容量を持つ下位階層の HDD がほとんど電源が止まっているために 50%以下の大幅な消費電力削減が期待される。しかし、データが停止した HDD にあるために読み出しにはディスク回転を立ち上げるための時間を確保するため予知が正確でなければデータアクセスに待ちが発生して読み出し速度性能が低下するリスクを持っている。

このリスクを定量的に把握するために待ち行列理論を用いて、アクセスされたデータをニアライン HDD から上位(オンライン) HDD にコピーする(ステージングする)時間と読み出し時間との確率統計的な解析を通じて、さまざまな条件でアクセスがミスヒットするエラーレートを計算した。

スーパーコンピュータのための大規模ストレージを想定して検討をした結果、ユーザからのジョブがある程度ジョブキューに存在していれば問題なく、読み書き速度の低下なく省電力性を改良できることが明らかになった。現在、キューにある待ちジョブが極端に少ない場合でもカバーできるようにより詳細な検討を続けている。

【高密度磁気記録の研究】

一方で、磁気記録の高密度化は大容量性と小型化及びビット単価の低廉化のために HDD としての開発以来の一貫した研究要素であるが、同時に必要な記録容量に対して HDD 台数の低下が可能という意味で省電力化にも大きな効果がある。研究内容としては高密度化であるが、本共同プロジェクト研究での目的は省電力化である。

今年度の成果は、コンピュータシミュレーションによる記録再生理論の検討、強磁界垂直磁気記録ヘッド、高密度信号処理方式、の 3 つに大別できる。

シミュレーションによる成果では、ディスク上のビットを微細エッチングによって分離形成するパターン媒体と大きな記録磁界を発生できる記録ヘッドを次々とずらしながら記録するシングル 2 次元記録の 2 つの代表的な次世代の高密度磁気記録方式について、その高記録密度性を検証した。パターン媒体では 5Gbit/inch² の高い記録密度が可能なが明らかになったが、そのためには記録媒体の磁気特性の均質性を極限まで高める必要がある。またシングル型 2 次元記録についても世界に先駆けてその検証に成功し、2Gbit/inch² の可能性を示した。このコンピュータシミュレーションによる検討は国際会議の招待講演に採択されている。

強磁界の記録ヘッドの研究については、秋田県高度技術研究所の提案の平面型垂直ヘッドの可能性を追求しているが、その主磁極構造を工夫することで強い磁界強度を保って優れた分解能を実現できるこ

とが有限要素法を用いたシミュレーションから判明し、それを具体的に実現する微細加工法も提示できた。

信号処理方式についても次世代の超高密度化に対応する低 SN 比の読み出し信号に適応可能な方式を検討しており、最近さまざまな分野で大きな成果をあげている LDPC 符号を用いる反復複合法によって垂直磁気記録の高密度化に高い性能が得られることが明らかになってきた。一方で、発生した誤りをリードソロモン符号による誤り訂正を行う直接的な手法も考えられ並行して検討している。

(3-2) 波及効果と発展性など

大規模の並列 HDD により構成されるストレージ技術は、インターネットや業務用データベースの大型情報ストレージサイトやスーパーコンピュータのストレージなどでは急速な拡大が続いており、今後の爆発的なネットワークの大規模化・高速化に加えて、スーパーハイビジョンなどの次世代高細精度の広帯域動画情報やそのアーカイブに代表される大規模情報の急速な展開を考えると極めて広範な ICT 技術での波及効果が想定され、将来のストレージシステムに関して先導的な開発が必要である。

実際、本共同プロジェクト研究は平成 19 年度より文部科学省の「次世代 IT 基盤構築のための研究開発」の一課題として採択され、「高機能・超低消費電力スピンドバイス・ストレージ基盤技術の開発」プロジェクトの取り組みの一部として検討を続けている。採択後も学術面での機敏な研究打合せの実施などこの大型プロジェクトを機動力のある形で支援する研究活動を担う重要な活動している。

[4] 成果資料

1. K. Fujimoto, N. Okada, K. Miura, H. Muraoka, H. Akaike, "Power-aware Storage-tiering Management for High-speed Tiered-storage Systems," 7th USENIX Conference on File and Storage Technologies, Feb 24-27, 2009
2. 赤池洋俊、藤本和久、岡田尚也、三浦健司、村岡裕明、「HPC 向け高速・大容量ストレージの省電力化を図る階層ストレージアーキテクチャと階層管理方式の提案」、情報処理学会全国大会、4A-1、2009
3. 岡田尚也、藤本和久、赤池洋俊、三浦健司、村岡裕明、「アクセス予測を利用した HPC 向け高速・大容量階層ストレージの階層管理方式における予測確率に関する検討」、情報処理学会全国大会、4A-2、2009
4. S. Okamoto, N. Kikuchi, T. Kato, O. Kitakami, K. Mitsuzuka, T. Shimatsu, H. Muraoka, H. Aoi, J.C. Lodder, "Magnetization behavior of nanomagnets for patterned

media application," J. Magn. Magn. Mater., 320, 22, Nov (2008) 2874-2879

5. S.J. Greaves, H. Muraoka, Y. Kanai, "Simulations of recording media for 1Tb/in²," J. Magn. Magn. Mater., 320, 22, Nov (2008) 2889-2893 [Invited]
6. Kenji Miura, Toshiyuki Ogawa, Hajime Aoi, Hiroaki Muraoka, Yoshihisa Nakamura, "New estimation methods of erase band width and track-edge noise in perpendicular magnetic recording," J. Magn. Magn. Mater., 320, 22, Nov (2008) 2908-2911
7. Y. Nakamura, M. Nishimura, Y. Okamoto, H. Osawa, H. Aoi, H. Muraoka, Y. Nakamura, "Iterative decoding algorithm using attenuated information in PMR with bit-patterned medium," J. Magn. Magn. Mater., 320, 22, Nov (2008) 3132-3135
8. Y. Okamoto, T. Bushisue, Y. Kurihara, Y. Nakamura, H. Osawa, H. Muraoka, Y. Nakamura, "Dependence of error patterns and BER performance on write pre-compensation in GPRML channels," J. Magn. Magn. Mater., 320, 22, Nov (2008) 3192-3195
9. N. Degawa, S. J. Greaves, H. Muraoka, Y. Kanai, "Characterization of a 2 Tbit/in² Patterned Media Recording System," IEEE Trans. Magn, 44, 11, pp. 3434-3437, Nov, 2008.
10. S.J. Greaves, Y. Kanai, H. Muraoka, "Magnetic Recording in Patterned Media at 5-10 Tb/in²," IEEE Trans. Magn, 44, 11, pp. 3430-3433, Nov 2008.
11. Y. Kanai, K. Hirasawa, T. Tsukamoto, K. Yoshida, S.J. Greaves, H. Muraoka, "Micromagnetic Recording Field Analysis of a Single-Pole-Type Head for 1-2 Tbit/in²," IEEE Trans. Magn, 44, 11, pp. 3609-3612, Nov 2008.
12. H. Osawa, M. Hino, N. Shinohara, Y. Okamoto, Y. Nakamura, H. Muraoka, "Simplified Neural Network Equalizer With Noise Whitening Function for GPRML System," IEEE Trans. Magn., Volume 44, Issue 11, Part 2, Nov. 2008 Page(s):3777 - 3780.
13. Y. Nakamura, M. Nishimura, Y. Okamoto, H. Osawa, H. Muraoka, "A New Burst Detection Scheme Using Parity Check Matrix of LDPC Code for Bit Flipping Burst-like Signal Degradation," IEEE Trans. Magn., Volume 44, Issue 11, Part 2, Nov. 2008 Page(s):3773 - 3776.
14. Y. Ohsawa, K. Yamakawa, H. Muraoka, "Writing Performance of a Planar Single-Pole Head With a Main Pole Fabricated by Ion-Beam Milling," IEEE Trans. Magn., Volume 44, Issue 11, Part 2, Nov. 2008 Page(s):3613 - 3616.